Name :- Aniket Kumar

Roll no :- PF 45

PRN no :- 1032171203

LAB ASSIGNMENT 03 [CCNLP]

Problem Statement :-

Watson Natural Language Understanding API -
Write a program that reads several news articles
or blogs online or from a text file and tells the
user what category the article is about. Alternati
-vely, the program classifies and sorts the article
and moves it to the proper category folder.

Objectives :

1. To study and explore IBM Watson Natural Language
Understanding API

2. To learn concept of text analysis and Sentim
-ents analysis and classification.

Theory :

- IBM Watson Natural Language Understanding APIs.

   It helps in analyzing text to extract metadata
   from content such as concepts, entities, keywords,
   categories, sentiment roles using natural language
   understanding. It can interpret text in 13
   different languages.

- Features of Text analysis.

Natural language Understanding includes a set of text analytics features that you can use to extract meaning from unstructured data.

These are :-

1] Categories :- Return a five-level taxonomy of the content.

2] Concepts :- Return high-level concepts in the content. For example, a research paper about deep learning might return the concept, "Artificial Intelligence" although the term is not mentioned.

3] Emotion :- Detects anger, disgust, fear, joy or sadness that is conveyed in the content or by the content around target phrases specified in the target parameter. We can analyze emotion for detected entities with entities emotion and for keywords with keywords emotion.

4] Entities :- Identifies people, cities, organization and other entities in the content.

5] Keywords :- Keywords of content are returned.

6] Metadata :- Returns information from the document, including author name, title, RSS/ATOM feeds, prominent page image, and publication data. Support URL and HTML input types only.

7.] **Relations :-** Recognizes when two entities are related and identifies the type of relation.

8.] **Semantic Roles:-** Parses sentences into subject, action, and object form.

9.] **Sentiment :-** Analyzes the general sentiment of your content of the sentiment towards specific target phrases.

10.] **Syntax:-** Return information about the tokens and sentences in the input text.

11.] **Summarization :-** Returns a summary of the source content, up to a maximum of three sentences.

• **Classification :-**

At the core of NLP lies text classification. Watson Natural language classifier (NLC) allows users to classify text in custom categories, at scale. Developer without a background in machine learning (ML) or NLP can enhance their applications using this service. NLC combines various advance ML techniques to provide the highest accuracy possible, without requiring a lot to training data.

Behind the scenes, NLC utilizes an ensemble of classification model, along with unsupervised and supervised learning techniques, to achieve its accuracy level. After your training data is

assembled : NLC evaluated your data against multiple Support vector Machine (SVMs) and a convolutional neural network (CNN) using IBM's Deep Learning As a Services (DLaaS)

Algorithm / Implementation :-

1. Create News Api access point and fetch news and store in dataframe

2. Use Natural Language Understanding Services using cloud ibm Credentials.

3. Use the Concept Service (which uses deep learning to define the concept of the entire news)

4. Create a column to store abstract information received from concept services.

5. Create a dataframe to store the Similar abstract information.

6. Create csv files of all different types of abstract information grouped by their Similar types

Platform :- 64-bit open source Linux, IBM cloud, Json

Input :- Text document, news article, online blogs.

Output :- Classification of input files in different categories

Conclusion :- Hence, learned the concepts of text analysis and sentiment analysis and classified analysis and sentiment analysis and classified the input document in respective category using IBM watson Natural language understanding API.

FAQ:-

1.] what is sentiment analysis?

Ans :- Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

2.] List different best classification algorithms used in NLP

Ans :- In NLP, the most use classification algorithm are :-

i) Naive Bayes

2) Support Vector Machines

3) Deep Learning : Convolutional Neural Network
                   Recurrent Neural Network